Iustum Aequum Salutare
XXI. 2025. 3. • 97–114.

ALGORITMIKUS ELFOGULTSÁG MINT JOGI DILEMMA

Lendvai Gergely Ferenc*

1. Bevezetés

A tanulmány bevezetéseként három példán keresztül szemléltetjük az algoritmikus elfogultság[1] megjelenését. A jelenség egyik legfontosabb pillanata közel tíz évvel ez előttre datálható. A COMPAS *(Correctional Offender Management Profiling for Alternative Sanctions[2])* algoritmus – amelyet az amerikai büntető igazságszolgáltatási rendszerben a mai napig használnak – egyik korai fejlesztése egy olyan mesterséges intelligencia (a továbbiakban: „MI") alapú eszköz volt, amelyet arra terveztek, hogy „megjósolja" a vádlottak visszaesésének valószínűségét, és ezzel támpontot adjon bírói döntésekhez és adatalapú szempontokkal szolgáljon a mérlegelés során.[3] A rendszer a maga nemében forradalmi volt; a szoftver fejlesztői ugyanis azt állították, hogy a COMPAS befolyásmentesen és objektíven, kizárólag statisztikai alapon, adatvezérelt módon mutat eredményeket. A valóság ezzel szemben kevésbé volt idealisztikus. A rendszert már egészen korán kritizálták szakértők és kutatók annak tévedései és olykor kiszámíthatatlan döntési mintázatai,[4] a szoftver módszertani hiányosságai[5] és

* PhD-hallgató, Pázmány Péter Katolikus Egyetem, Jog- és Államtudományi Kar.
ORCID: https://orcid.org/0000-0003-3298-8087

1 A szövegben következetesen az algoritmikus elfogultság kifejezést használjuk, mivel tartalmilag ez fedi a legjobban az angol *„algorithmic bias"* jelentését.

2 Magyarul talán „Büntetés-végrehajtási elkövetői menedzsment profilalkotás az alternatív szankciókhoz" lenne a megfelelő fordítás.

3 Christoph Engel – Lorenz Linhardt – Marcel Schubert: Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism. *Artificial Intelligence and Law,* Vol. 33. (2024) 3–4. https://doi.org/10.1007/s10506-024-09389-8

4 Tim Brennan – William Dieterich – Beate Ehret: Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System. *Criminal Justice and Behavior,* Vol. 36., No. 1. (2008) https://doi.org/10.1177/0093854808326545

5 Jay P. Singh: Predictive Validity Performance Indicators in Violence Risk Assessment: A Methodological Primer. *Behavioral Sciences & the Law,* Vol. 31., No. 1. (2013) 8–12. https://doi.org/10.1002/bsl.2052

https://doi.org/10.69695/ias.2025.3.07

etikai dilemmái miatt.[6] 2016-ban viszont az algoritmikus elfogultság rövid, de igen gazdag történetének egyik legjelentősebb ügyével került a köztudatba a COMPAS. A ProPublica független, non-profit, főleg oknyomozó újságírásra szakosodott szervezet széleskörű vizsgálatot folytatott a szoftver által javasolt döntések kapcsán és azt állította, hogy az algoritmus jelentős faji elfogultságot mutat az afro-amerikai emberek hátrányára.[7] A szervezet kutatói és újságírói a floridai Broward megyében több mint 10000 vádlott adatait vizsgálta meg, kifejezetten a lehetséges faji elfogultságra összpontosítva. A Broward megyei seriff hivatalától két év COMPAS-pontszámát szerezték meg, amely a 2013-ban és 2014-ben „pontozott" összes személyt felöleli. A hatalmas adatmennyiséget magában foglaló vizsgálat végén a ProPublica kiszámította az algoritmus általános pontosságát, és feltárta, hogy a nem visszaeső afro-amerikai vádlottakat majdnem kétszer nagyobb valószínűséggel minősítették tévesen magas kockázatúnak, mint fehér társaikat (45% vs. 23%), és a visszaeső fehér vádlottakat gyakrabban minősítették tévesen alacsony kockázatúnak, mint az afro-amerikai visszaesőket (48% vs. 28%). Az erőszakos visszaesés-elemzés még drasztikusabb képet vázolt fel: a korábbi bűncselekmények, a jövőbeni visszaesés, az életkor és a nem figyelembevételével a COMPAS 77%-kal nagyobb valószínűséggel adott magasabb kockázati pontszámot az afro-amerikai vádlottaknak, mint a fehér vádlottaknak. Bár a ProPublica elemzését Flores és társai módszertani hiányosságokra és hibás adatokra tekintettel sok helyütt megkérdőjelezték,[8] a COMPAS és az algoritmus faji alapú megkülönböztetésre okot adó gyakorlatáról a mai napig születnek empirikus kutatások, javarészt megerősítve a ProPublica állításait.[9]

A második, talán a mai napig az egyik leghíresebb esete az algoritmikus elfogultságnak az Amazon MI-alapú toborzó alkalmazása és rövid „élete". A BBC által „szexista MI"-nek[10] titulált szoftvert 2014-ben kezdte el fejleszteni az Amazon azzal

---

6 Georgios Bouchagiar: Is Europe prepared for Risk Assessment Technologies in criminal justice? Lessons from the US experience. *New Journal of European Criminal Law,* Vol. 15., No. 1. (2024) https://doi.org/10.1177/20322844241228676

7 Julian Angwin – Jeff Larson – Surya Mattu – Lauren Kirchner: Machine Bias. *ProPublica,* 2016. május 23. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing A tanulmányhoz használt módszertanhoz ld. Jeff Larson – Surya Mattu – Lauren Kirchner – Julia Angwin: How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica,* 2016. május 23. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

8 Anthony W. Flores – Kristin Bechtel – Christopher T. Lowenkamp: False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used across the Country to Predict Future Criminals and It's Biased against Blacks.' *Federal Probation,* Vol. 80., No. 2. (2016) 44–45. Ugyancsak ld. Cynthia Rudin – Caroline Wang – Beau Coker: The Age of Secrecy and Unfairness in Recidivism Prediction. *Harvard Data Science Review,* Vol. 2., No. 1. (2020) https://doi.org/10.1162/99608f92.6ed64b30

9 Riccardo Fogliato – Alexandra Chouldechova, – Max G'Sell: Fairness Evaluation in Presence of Biased Noisy Labels. *International Conference on Artificial Intelligence and Statistics,* 2020. június 3. (2020) 2325–2336.; Francesca Lagioia – Riccardo Rovatti – Giovanni Sartor: Algorithmic fairness through group parities? The case of COMPAS-SAPMOC. *AI & Society,* Vol. 38., No. 2. (2022) 463–464. https://doi.org/10.1007/s00146-022-01441-y

10 Amazon scrapped ‚sexist AI' tool. *BBC,* 2018. október 10. https://www.bbc.com/news/technology-45809919

a céllal, hogy a leendő alkalmazottak felvételi folyamatát egyszerűsítések. Maga a mechanizmus tipikusan olyan feladatokat érintett, amelyek a humánerőforrás területén dolgozók munkakörét fedik le: önéletrajzokat elemzett és szortírozott, illetve a különböző pozíciókhoz szükséges legjobb jelentkezőket igyekezett azonosítani az önéletrajzban feltüntetett információk alapján. Az eszközt sajátos módon az Amazon a saját, egy évtized alatt a vállalathoz benyújtott önéletrajzokon képezte ki, és gépi tanulási algoritmusokat használtak a minták azonosítására és a jelöltek értékelésére. A program egy specifikus, egytől öt csillagig terjedő pontozást vezetett be.[11] 2018-ra nyilvánvalóvá vált, hogy az eszköz szisztematikusan diszkriminálja a nőket, különösen a műszaki szerepkörökben, amely jelentős hibát fedett fel a mögöttes adatokban és algoritmusokban.[12] Az előítéletesség gyökere a képzési adatokban rejlett: a korábbi önéletrajzok ugyanis túlnyomórészt férfi jelentkezőktől származtak és az algoritmus megtanulta, hogy a sikeres, felvett és alkalmazandó jelöltekhez a férfiakat társítsa.[13]

Végül egy egészen friss, nagy médiavisszhangot generáló eseménnyel érdemes zárni a példák sorát. Az MI eddig soha nem látott demokratizálódása során, amelyben jelentős szerepet játszottak a különböző nagy nyelvi modellek elterjedése (például ChatGPT), heves verseny alakult ki a képgeneráló MI-rendszerek fejlesztése vonatkozásában, amelynek 2024-ben a legprominensebb résztvevői a DALL-E (ma már a ChatGPT-be integrált) és a MidJourney nevű modellek voltak. Új versenyzőként a Google kifejlesztette a Gemini (Bard) névre hallgató rendszerét, amelynek egyik funkciója a képgenerálás volt.[14] Ez utóbbi funkció bemutatkozása ugyanakkor igen ellentmondásosnak bizonyult. Ellentétben ugyanis a fenti két példával, amelyek esetében az az adatkészletekben rejlő „inherens" társadalmi alapú megkülönböztetés volt a polémiák alapja, a Gemini esetében pont az ilyen jellegű problémák kiszűrésére irányuló „túlkompenzálás" volt az elfogultság jelenségének eredője. A Gemini ugyanis többször is történelmileg pontatlan és etnikailag sokszínű ábrázolásokat készített fehér vagy hagyományosan fehérnek ábrázolt emberekről és történelmi szereplőkről. Ennek két eklatáns példája az egészen bizarr hatást keltő afro-amerikai nemzetszocialista katona (bal) és az amerikai indián *(amerind)* viking harcos (jobb). (1. ábra)

---

[11] Az Amazonon árult termékek értékeléséhez volt hasonló a pontozási rendszer.

[12] Lennart Hofeditz – Milad Mirbabaie – Audrey Luther – Riccarda Mauth – Ina Rentemeister: Ethics Guidelines for Using AI-based Algorithms in Recruiting: Learnings from a Systematic Literature Review. *Proceedings of the ... Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences,* 2022. január 1. (2022) 145. https://doi.org/10.24251/hicss.2022.018

[13] Jeremy Hsu: Can AI hiring systems be made antiracist? Makers and users of AI-assisted recruiting software reexamine the tools' development and how they're used. *IEEE Spectrum,* Vol. 57., No. 9. (2020) 9. https://doi.org/10.1109/mspec.2020.9173891

[14] Vö. Muhammad Imran – Norah Almusharraf: Google Gemini as a next generation AI educational tool: a review of emerging educational technology. *Smart Learning Environments,* Vol. 11., No. 1. (2024) https://doi.org/10.1186/s40561-024-00310-z ; Hamid Reza Saeidnia: Welcome to the Gemini era: Google DeepMind and the information industry. *Library Hi Tech News,* December 26, 2023. https://doi.org/10.1108/lhtn-12-2023-0214

1. ábra. A Gemini által generált képek (Forrás: Matthew Field, Telegraph, 2024[15]*)*

A Gemini kimeneteinek elfogultsága a korábbi két esethez hasonlóan az MI képzéséhez használt adatokban gyökerezik, amelyek bizonyos nézőpontokat és kérdéseket, mint például az inkluzivitást felül-, míg másokat, így a történelmileg hű ábrázolást alulreprezentálták, ami bár nem szándékolt, de kétséget kizárólag egyaránt szürreális és súlyosan sértő eredményekhez vezettek. Bár a Google azonnal reagált és elnézést kért, illetve rendkívüli gyorsasággal kigyomlálta a problémára okot adó elemeket,[16] a Gemini-t a mai napig kritizálják etnikai elfogultsága miatt.[17]

E három példa hűen tükrözi Cathy O'Neil, az algoritmikus elfogultság egyik legelismertebb kutatójának gondolatát, aki az elsők között fejtette ki, hogy az új technológiákkal nem megszüntetjük, hanem csupán álcázzuk az emberi elfogultságot.[18] Bár az adatok kétségtelenül, természetükből adódóan függetlennek és objektíveknek tűnhetnek, az adatokkal dolgozó egyén viszont nem az – éppen ebből ered az elfogultság. Jelen tanulmány célja, hogy az algoritmikus elfogultságot konceptualizálja és a jelenlegi jogi válaszokat bemutassa. E körben elsősorban az európai jogalkotásra fókuszálunk

---

15 Matthew Field: From Black Nazis to female Popes and American Indian Vikings: How AI went 'woke'. *The Telegraph,* 2024. február 23. https://tinyurl.com/2uke3bt9

16 Ali Robertson: Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis. *The Verge,* 2024. február 21. https://tinyurl.com/mfsvs2xc

17 Julia Barroso Da Silveira – Ellen Alves Lima: Racial Biases in AIs and Gemini's Inability to Write Narratives About Black People. *Emerging Media,* Vol. 2., No. 2. (2024) https://doi.org/10.1177/27523543241277564

18 Cathy O'Neil: *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London, Penguin, 2016. 19.

és igyekszünk feltárni olyan hiányosságokat, amelyeket érdemes a jövő jogalkotóinak orvosolni. A tanulmány jogi elemzésre és szakirodalmi áttekintésre épül. Bár az algoritmikus elfogultság hazai, magyar nyelvű irodalma igen kezdetleges, reméljük, hogy jelen dolgozat vitaindító és inspiráló lehet a digitális technológiák és a mesterséges intelligencia szabályozása iránt érdeklődő kutatók számára.

2. Az algoritmikus elfogultság fogalma

Akárcsak a digitális jelenségek javarészében, az algoritmikus elfogultság esetében is problémás az egységes fogalom megalkotása.[19] Általánosságban véve az algoritmikus elfogultság alatt olyan MI-rendszerekben vagy MI-alapú rendszerekben előforduló szisztematikus és strukturált hibákat és elfogultsági pontokat értünk, amelyek részrehajló[20] eredményeket és egyenlőtlenségeket eredményeznek anélkül, hogy erre valamilyen módon igazolható indok lenne.[21] Javasolt a fogalmat elemeire bontani. Az algoritmikus elfogultság és az MI kapcsolata talán magától értetődőnek hat, ugyanakkor érdemes röviden kitérni ennek a hátterére. Ahogy Shin és Shin is aláhúzta, az emberi kognitív előítéletek sokszor „bekerülnek" az algoritmusokba, az MI pedig képes felerősíteni ezeket.[22] Ahogy a szerzők fogalmaznak: „ahhoz, hogy az MI igazodjon az emberek által kívánatosnak tartott funkciókhoz és feladatokhoz, meg kell tanulnia ezeket a preferenciákat, viszont az emberi értékek megtanulása kockázatokat hordoz magában."[23] Az elfogultság és az MI kapcsolata tehát technológiai szempontból inherens.[24] Az algoritmikus elfogultság fogalmának központi eleme a szisztematikus, rendszerszintű elfogultság. Bár a rendszerszintűség és az elfogultság irodalma igen gazdag,[25] jelen kontextusban két definíciós elemre érdemes kitérni. Egyfelől, a szisztematikus jelleg alatt nem egyedi, egyszeri hibából adódó esetekre gondolunk.

---

19 Jin-Young Kim – Sung-Bae Cho: An information theoretic approach to reducing algorithmic bias for machine learning. *Neurocomputing,* Vol. 500. (2022) 26–27. https://doi.org/10.1016/j.neucom.2021.09.081

20 A szakirodalom az angol *„unfair"* szóval él, ugyanakkor az „igazságtalan" fordítás fogalmilag nem érintené az algoritmikus elfogultság elfogultságra, vagy akár bizonyos értelemben méltánytalanságra utaló jellegét. A fogalmi kérdésekhez lásd a *„fairness"* és *„justice"* distinkciót itt: Lionel P. Robert – Casey Pierce – Liz Marquis – Sangmi Kim – Rasha Alahmad: Designing fair AI for managing employees in organizations: a review, critique, and design agenda. *Human-Computer Interaction,* Vol. 35., No. 5–6. (2020) https://doi.org/10.1080/07370024.2020.1735391

21 Nima Kordzadeh – Maryam Ghasemaghaei: Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems,* Vol. 31., No. 3. (2021) 388. https://doi.org/10.1080/0960085x.2021.1927212

22 Donghee Shin – Emily Y. Shin: Data's Impact on Algorithmic Bias. *Computer,* Vol. 56., No. 6. (2023) 90. https://doi.org/10.1109/mc.2023.3262909

23 Uo.

24 Michael Veale – Max Van Kleek – Reuben Binns: Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada,* February 4, 2018. (2018) https://doi.org/10.31235/osf.io/8kvf4

25 Ehhez ld. Gabbrielle M. Johnson: Algorithmic bias: on the implicit biases of social technology. *Synthese,* Vol. 198., No. 10. (2020) https://doi.org/10.1007/s11229-020-02696-y

Visszatérve a Gemini példájára, a képgeneráló rendszerek gyakran tévesztenek, vagy állítanak elő nem megfelelő kimenetet, a Google fejlesztése esetében viszont egy folyamatosan, visszatérő módon előforduló problémáról volt szó, amely nem egy egyedi *„glitch"*, hanem – *ab ovo* – a modell kialakítottságának sajátja volt. Másfelől, a rendszerszintűség egyet jelent a reprodukálhatósággal is,[26] amely leginkább az adatgyűjtés és -feldolgozás során ejtett hibákból és a megfelelő transzparencia-mechanizmusok hiányából,[27] illetve – éppen a Gemini esetében – az adatkészletben rejlő „adott" hibák „túlbuzgó" emberi tevékenységből adódik. Talán a legnehezebben megragadható fogalmi elempár a részrehajló eredmények és egyenlőtlenségek, mint kimenetek. Ennek oka sokrétű. Egyrészről a részrehajló „jelleg" és az egyenlőtlenség korántsem technológiai fogalmak. Ahogy Kordzadeh és Ghasemaghaei összefoglaló tanulmányából is kiderül, a részrehajlásnak és „elfogultságnak" igen gazdag „techno-filozófiai" szakirodalma van.[28] Ha bölcseleti oldalról vizsgáljuk a részrehajlás kérdését, akkor a szerzők szintézisére hivatkozva az elfogultság a következő jelenségeket takarja:

1. egy algoritmus egyenlőtlenül osztja el az előnyöket és hátrányokat a különböző egyének vagy csoportok között;
2. ez az egyenlőtlen elosztás az egyének eredendő tulajdonságaiban, tehetségében vagy szerencséjében rejlő különbségekre vezethető vissza, így
3. az algoritmikus elfogultság egy algoritmus kimeneteiben megjelenő, az egyenlőségtől való szisztematikus eltérés.[29]

Másrészről az egyenlőtlenség mint probléma lényegében implikálja, hogy az algoritmusok kialakításakor létezik egy etikai standard, amelynek alapja, hogy az algoritmus alapú „egyenlőség" elérhető és megvalósítható, ha a nem szándékosan beépített torzításokat és részrehajlásokat kiszűrjük. Ezt egyes kutatók „észlelt pártatlanságnak" *(perceived fairness)* nevezik, amely utal az algoritmusok által generált kimenetek, a döntéshozatal és az algoritmusok kialakításának pártatlan, azaz nem megkülönböztető jellegére.[30]

---

[26] Zhisheng Chen: Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications,* Vol. 10., No. 1. (2023) 3. https://doi.org/10.1057/s41599-023-02079-x

[27] Jonathan Dodge – Q. Vera Liao – Yunfeng Zhang – Rachel K. E. Bellamy – Casey Dugan: Explaining models: an empirical study of how explanations impact fairness judgment. *Proceedings of the 24th International Conference on Intelligent User Interfaces. Marina Del Ray, CA, USA*., March 17, 2019. (2019) 275–285. https://doi.org/10.1145/3301275.3302310

[28] Kordzadeh–Ghasemaghaei i. m. 7–8.

[29] Uo.; Sahil Verma – Julia Rubin: Fairness definitions explained. *2018 IEEE/ACM International Workshop on Software Fairness (FairWare), Gothenburg, Sweden*., May 29, 2018. (2018) https://doi.org/10.1145/3194770.3194776; Reuben Binns Max Van Kleek – Michael Veale – Ulrik Lyngs – Jun Zhao – Nigel Shadbolt: 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. *arXiv (Cornell University),* April 21, 2018. (2018) 377. http://arxiv.org/abs/1801.10408

[30] Binns et al. i. m.; Kordzadeh–Ghasemaghaei i. m. 9–10.

3. Az algoritmikus elfogultság és a szabályozás kapcsolata, illetve az amerikai perspektívák

A jogi szabályozás az algoritmikus elfogultság esetében polemikus: kérdés tucatjával érkezik a szabályozó felé, a válaszokra viszont sokszor éveket kell várni, sőt, ha mégis sikerül megoldást találni, talán már okafogyottá is válik az eredeti kérdés.[31] Jelen szegmensben az algoritmikus elfogultság szabályozási dilemmáit mutatjuk be, azaz, hogy egyáltalán *miért* jogi dilemma az algoritmikus elfogultság kezelése.

Az algoritmikus elfogultság elsősorban adatvédelmi és a mesterséges intelligencia általános szabályozását érinti. A „szabályozás" e körben nem specifikus, kifejezetten az elfogultságra vonatkozó jogszabályokat, hanem bizonyos nagyobb tárgykörök vonatkozó szabályait érinti, amelyek alkalmazhatók az algoritmikus elfogultságra. Nem túlzás azt állítani, hogy a jelenséget érintő szabályozási keretet a nemzeti, regionális és globális szintű kezdeményezések sokasága jellemzi, nem is beszélve a platformok és a mesterséges intelligenciával foglalkozó cégek saját szabályairól. Fejlődő, az MI-szabályozás alszegmensének tekinthető szabályozási környezetről beszélhetünk tehát, amelyet leginkább a harmonizáció alacsony szintje és a töredezettség jellemez.[32] Ahogy más mesterséges intelligenciát érintő kérdésekben,[33] az algoritmikus elfogultság esetén is alapvetően két fő szabályozási irány azonosítható: az európai, erősen korlátozó, ugyanakkor felhasználó-központú és az amerikai, fragmentáltabb, „kevert" jellegű, a technológiai és gazdasági fejlődést is támogató attitűd.[34] Tekintve, hogy az európai szabályozást a következő fejezetben részletezem, a következőkben az amerikai megközelítést ismertetem röviden.

Wang és társai szerint az amerikai jogi keretrendszer az algoritmikus elfogultság tekintetében az alapvető polgárjogi védelemben és a tizennegyedik módosításban gyökerezik, kiemelt hangsúllyal három alapvető elvre: az egyenlőségre, a megkülönböztetésmentességre és az átláthatóságra.[35] A fentebb is említett „kevert" szabályozási rendszer abból fakad, hogy az algoritmikus elfogultságra egyaránt irányadók a vonatkozó, elsősorban munkakereséssel, hitelfelvétellel és egy állásbetöltéssel kapcsolatos törvények, mint a *Fair Credit Reporting Act* és az *Equal Employment Opportunity Commission,* amelyek többek között kitérnek az algoritmusok pártatlanságának fontosságára is,[36] ezen túlmenően pedig a bírósági ítéletek is jelentős szere-

---

31 Gergely Ferenc Lendvai: Sharenting as a regulatory paradox – a comprehensive overview of the conceptualization and regulation of sharenting. *International Journal of Law Policy and the Family,* Vol. 38., No. 1. (2024) 17. https://doi.org/10.1093/lawfam/ebae013

32 Xukang Wang – Ying Cheg Wu – Xueliang Ji – Hongpeng Fu: Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices. *Frontiers in Artificial Intelligence,* Vol. 7. (2024) 4–6. https://doi.org/10.3389/frai.2024.1320277

33 Vö. Gosztonyi Gergely – Lendvai Gergely: Deepfake és dezinformáció. Mit tehet a jog a mélyhamisítással készített álhírek ellen? *Médiakutató,* Vol. 25., No. 1. (2024) https://doi.org/10.55395/mk.2024.1.3

34 Wang et al. i. m. 4–6

35 Uo.

36 Mark MacCarthy: Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms. *Cumberland Law Review,* Vol. 48., No. 1. (2017) 74–89.

pet játszanak mind az e törvények értelmezésében az algoritmikus diszkriminációval kapcsolatos esetekben, mind az egyre gyakrabban előforduló, automatizált rendszerek által okozott foglalkoztatási és lakhatási előítéletességgel foglalkozó ügyekben.[37]

Átfogó szövetségi törvény, amely kifejezetten az algoritmikus elfogultság kezelésére szolgálna viszont nincs, még annak ellenére sem, hogy az e kézirat írásakor leköszönő Joe Biden elnök 2023 októberében kiadott végrehajtási rendeletében *(executive order)* a biztonságos, megbízható mesterséges intelligenciára vonatkozóan külön kiemelésre került, hogy az amerikai szabályozás kifejezetten ügyelni fog az algoritmikus elfogultság megfelelő kezelésére.[38] A dokumentum erre több javaslatot is tesz:

1. Egyértelmű iránymutatást kell adni a bérbeadók, a szövetségi juttatási programok és a szövetségi vállalkozók számára annak megakadályozása érdekében, hogy a mesterséges intelligencia algoritmusokat a megkülönböztetés súlyosbítására használják.
2. Az Igazságügyi Minisztérium és a szövetségi polgárjogi hivatalok közötti együttműködést biztosítani és támogatni kell, hogy az MI-vel és kifejezetten az algoritmikus elfogultsággal érintett polgári jogi jogsértéseket megfelelően ki lehessen vizsgálni.
3. A büntető igazságszolgáltatási rendszerben biztosítani kell az algoritmikus pártatlanságot és méltányosság.
4. A 2) és 3) pontok esetében ki kell alakítani „legjobb gyakorlatokat".[39]

Bár a Biden-adminisztráció ambiciózusnak tűnt 2023 őszén, amelyet megelőzött egy évvel korábban a *„Blueprint for an AI Bill of Rights"* javaslat,[40] az amerikai MI-szabályozás jövője a Trump-adminisztráció alatt teljes fordulatot vett. Lényegében pár órával beiktatása után ugyanis a republikánus elnök visszavonta a Biden-féle végrehajtási rendeletet, sőt, a Fehér Ház kommünikéjében „ártalmasnak" nyilvánította azt,[41] egy későbbi kiadványban pedig kifejezetten aláhúzta, hogy semmiféle „fejlesztés-korlátozó" szabályozást nem kíván bevezetni a mesterséges intelligencia tekintetében.[42] Ahogy azt Saed is kifejti, a Trump-adminisztráció radikálisan liberális MI-(nem) szabályozásának következményei hosszú-, de még rövidtávon is beláthatatlanok, többek között amiatt is, mert az MI-fejlesztéssel járó kockázatok felmérésére nem került

37 Wang et al. i. m. 4.

38 Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. *WH.Gov*, 2023. október 30. (A kézirat leadásakor elektronikusan elérhető volt, de a Trump-adminisztráció azóta törölte.)

39 Uo.

40 Blueprint for an AI Bill of Rights. *WH.Gov*, 2022–2023. (A kézirat leadásakor elektronikusan elérhető volt, de a Trump-adminisztráció azóta törölte.)

41 Initial Rescissions of Harmful Executive Orders and Actions. White House, 2025. január 20. https://tinyurl.com/4wuhbe8h

42 Removing Barriers to American Leadership in Artificial Intelligence. White House, 2025. január 23. https://tinyurl.com/2u6zf73x

különösebb hangsúly eddig.[43] Azt viszont egészen biztosan állíthatjuk, hogy a Biden-adminisztráció által felvázolt progresszív, az emberek biztonságát előtérbe helyező MI-szabályozás visszatérte belátható időn belül nem valószínű. Bár a kutatások e téren egyelőre spekulatívak, Novelli és társainak tanulmánya fontos megállapításokat tesz. A szerzők két lehetséges irányt vázolnak fel az AI-szabályozás terén a Trump-adminisztráció alatt: az egyik a decentralizált, állami szintű szabályozás, ahol az egyes államok saját törvényeket alkotnak, a másik pedig a „szövetségi szabályozás térnyerése", amelynek célja az egységes, dereguláción alapuló, innovációbarát szabályozási környezet megteremtése.[44] Mindkét modell viszont komoly jogi és politikai kihívásokat vet fel, legyen szó a fragmentált szabályozás kockázatáról, vagy éppen a potenciális alkotmányjogi konfliktusokról.

4. Az uniós jog és az algoritmikus elfogultság ezer arca

Az algoritmikus elfogultság kapcsán – jelen dolgozat írásakor – egyértelműen az európai jogi keretrendszer nyújtja a legelőremutatóbb szabályozási megközelítést, még akkor is, ha egyesek hevesen kritizálják az uniós jogot amiatt, mert nem tér ki minden potenciális kockázatra.[45] Több oldalról érdemes megközelíteni az uniós szabályozást, a következőkben specifikus esetkörökre bontva ismertetjük a jogszabályok adta lehetőségeket.

Először érdemes az általános, ugyanakkor kifejezetten algoritmus-specifikációtól mentes keretrendszert ismertetni. Ehhez elsőként az adatvédelmi kérdéseket és az Általános adatvédelmi rendelet (GDPR)[46] megoldásait mutatjuk be. A GDPR 22. cikke e körben kulcsfontosságú. E cikk (1) bekezdése szerint ugyanis az „érintett jogosult arra, hogy ne terjedjen ki rá az olyan, kizárólag automatizált adatkezelésen – ideértve a profilalkotást is – alapuló döntés hatálya, amely rá nézve joghatással járna vagy őt hasonlóképpen jelentős mértékben érintené."[47] E rendelkezés alapja az a preambulumban is megfogalmazott cél, hogy a személyes adatok kezelése az embereket kell, hogy szolgálja,[48] illetve hogy biztosított legyen a tisztességes, és legfőképp átlátható adatkezelés.[49] A rendelkezés fő kérdése, különösen az algoritmikus elfogultság kapcsán

---

43 Ferial Saed: The Uncertain Future of AI Regulation in a Second Trump Term. *Stimson,* 2025. március 13. https://tinyurl.com/bnuzdttt

44 Claudio Novelli – Akriti Gaur – Luciano Floridi: Two Futures of AI Regulation under the Trump Administration. (Preprint.) *SSRN,* 2025. március 30. http://dx.doi.org/10.2139/ssrn.5198926

45 Ehhez ld. Philipp Hacker: Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review,* Vol. 55., No. 4. (2018) 1143–1146. https://doi.org/10.54648/cola2018095

46 Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet) (EGT-vonatkozású szöveg) HL L 119., 04/05/2016,1–88. o.

47 GDPR 22. cikk (1) bekezdés.

48 GDPR 4. preambulumbekezdés.

49 GDPR 60. preambulumbekezdés.

a joghatás, illetve a „hasonlóképpen jelentős mértékű" más fajta döntés definiálása, amelyre a GDPR csak absztraháltan vállalkozik a preambulumban.[50] A joghatás alatt elsősorban olyan döntéseket vagy tevékenységeket értünk, amely érinti valakinek a jogait, jogi státuszát, vagy akár szerződéses jogait. Barbosa és Félix szerint ilyenre lehet példa a szavazati jogra, a fogyatékosság miatti havi nyugdíjra való jogosultságra, vagy egy országba való belépési képességre gyakorolt hatások.[51] A hasonlóképpen jelentős mértékű hatással járó döntés interpretálása egy fokkal viszont komplexebb.[52] Bár a 71. preambulumbekezdés ad némi támaszt,[53] példálózó eseteken túlmenően nem derül ki a GDPR szövegéből mi a pontos köre az ilyen típusú döntéseknek.[54] Barbosa és Félix e helyütt az javasolja, hogy a rendelkezést elsősorban annak súlya, mintsem automatikus jellege szerint kell értékelni; azaz, bár kétségtelenül automatizált módon ajánl egy zeneszámot a streaming-platform, feltételezhetően ez nem jár szignifikáns, vagy akár joghatással összehasonlítható hatással.[55] Fontos megjegyezni viszont, hogy vannak kivételek a 22. cikk (1) bekezdése alól, így ha *A)* szerződéskötés vagy szerződés teljesítés miatt szükséges az adatkezelés, *B)* az uniós jog lehetőséget ad rá, illetve ha *C)* az érintett kifejezetten hozzájárul, nem szükséges a fenti rendelkezést alkalmazni.[56] Az A) és C) esetkörnél viszont fontos kitételt említ meg a Rendelet, mivel „az adatkezelő köteles megfelelő intézkedéseket tenni az érintett jogainak, szabadságainak és jogos érdekeinek védelme érdekében, ideértve az érintettnek legalább azt a jogát, hogy az adatkezelő részéről emberi beavatkozást kérjen, álláspontját kifejezze, és a döntéssel szemben kifogást nyújtson be."[57] Ahogy Bygrave is jelzi, ezek a lehetőségek ugyan alapvetően *ex post* jellegűek, mégis rendkívül előremutatóak a felhasználó- és az adatvédelem területén, és jelentős garanciákat hordoznak egy potenciális, sértő algoritmikus elfogultság eredményeként megszületett döntéssel szemben.[58]

---

50 GDPR 70. preambulumbekezdés.

51 Sandra Barbosa – Sara Félix: Algorithms and the GDPR: An analysis of article 22. *Anuário da Proteção de Dados,* 2021. 75.

52 Uo. 75–76.

53 […] természetes személyekre vonatkozó személyes jellemzők bármilyen automatizált személyes adatok kezelése keretében történő kiértékelése, különösen az érintett munkahelyi teljesítményére, gazdasági helyzetére, egészségi állapotára, személyes preferenciáira vagy érdeklődési körökre, megbízhatóságra vagy viselkedésre, tartózkodási helyére vagy mozgására vonatkozó jellemzők elemzésére és előrejelzésére.

54 Lee A. Bygrave: Article 22. In: Christopher Kuner –Lee A Bygrave – Christopher Docksey – Laura Drechsler (szerk.): *The EU General Data Protection Regulation (GDPR): A Commentary*. Oxford, Oxford University Press, 536.

55 Barbosa–Félix i. m. 76.

56 GDPR 22. cikk (2) bekezdés.

57 GDPR 22. cikk (3) bekezdés.

58 Bygrave i. m. 538.

Bár a DSA, azaz a digitális szolgáltatásokról szóló rendelet[59] elsősorban a platformszabályozás területén vezet be úttörő szabályokat,[60] részben lehet alkalmazni a rendelkezéseit az algoritmikus elfogultságra is. A DSA – Husovec szavaival élve – „megkülönböztető" szabályozást alkalmaz, amelynek értelmében a különböző platformok azok kategóriája és mérete alapján vezet be szabályokat.[61] Az algoritmusok kérdése két helyütt kerül elő hangsúlyosan a DSA-ban. Elsőként a 14. cikkben kerül említésre, amely a szerződési feltételekről rendelkezik, és külön kitér arra, hogy az algoritmikus döntéshozatalról szóló információkat a közvetítő szolgáltatóknak „világosan, egyszerűen, érthetően, felhasználóbarát módon" kell megfogalmazniuk és könnyen hozzáférhetővé kell tenniük az adott szolgáltatást igénybe vevő számára.[62] Jelentősebb fókusz kerül az algoritmusokra és azok átláthatóságára az online óriásplatformok és nagyon népszerű online keresőprogramok esetében.[63] Ezek olyan platformok, amelyek „havonta átlagosan legalább 45 millió, a szolgáltatást aktívan igénybe vevővel rendelkeznek az Unióban."[64] E helyütt megjegyzendő viszont, hogy konjunktív feltétele az online óriásplatform státusznak az is, hogy az Európai Unió Bizottsága annak is minősítse az adott platformot.[65] A minősítési eljárás nem egyszeri, és a Bizottság folyamatosan ellenőrzi és bővíti a VLOPSE-k listáját,[66] hogy a lehető legteljesebb körű képet lehessen kapni arról, kik is a platformok „óriásai". A VLOPSE-kra számos új, igen progresszív szabály vonatkozik; az algoritmusok kérdéskörét tekintve viszont a kockázatértékelés kötelezettsége a legszignifikánsabb új szabály. A DSA 34. cikk értelmében ugyanis a VLOPSE-knak „gondosan azonosítaniuk, elemezniük és értékelniük kell a szolgáltatásaik és a kapcsolódó rendszerek – többek között az algoritmikus rendszerek – kialakításából vagy működéséből vagy a szolgáltatásaik használatából eredő rendszerszintű kockázatokat".[67] Ilyen rendszerszintű kockázatok közé tartozik többek között a jogellenes tartalmak megosztása, az alapvető jogokat érő negatív hatások és gyakorlatok, a demokratikus diskurzust és a közbiztonságot érintő veszélyek, illetve nemi alapú erőszakkal, a közegészség és a kiskorúak védelmével összefüggő, továbbá a személyek testi és szellemi jóllétére gyakorolt negatív következmények.[68] Az algoritmikus elfogultság e rendszerszintű kockázatok mindegyik formájában manifesztálódhat – gon-

---

59 Az Európai Parlament és a Tanács (EU) 2022/2065 rendelete (2022. október 19.) a digitális szolgáltatások egységes piacáról és a 2000/31/EK irányelv módosításáról (digitális szolgáltatásokról szóló rendelet) (EGT-vonatkozású szöveg). PE/30/2022/REV/1; HL L 277., 2022.10.27, 1–102. o.

60 Gergely Ferenc Lendvai: Taming the Titans? – Digital Constitutionalism and the Digital Services Act. *ESSACHESS,* Vol. 17., No. 34. (2024) 169–172. http://dx.doi.org/10.21409/0M3P-A614

61 Martin Husovec: The DSA's Scope Briefly Explained. *SSRN,* July 4, 2023. 1–2.

62 Ehhez még ld. DSA 45. preambulumbekezdés.

63 E két platformfajtát legtöbbször „VLOPSE" néven rövidítjük, utalva az angol elnevezésre *(Very Large Online Platforms and Search Engines).*

64 DSA 33. cikk (1) rendelkezés.

65 Uo., és DSA 33. cikk (4) bekezdés.

66 Supervision of the designated very large online platforms and search engines under DSA. European Commission, 2024. december 17. https://tinyurl.com/376xzn37

67 DSA 34. cikk (1) bekezdés.

68 Uo.

doljunk csak egyes vélemények szisztematikus vagy olykor „láthatatlan" törlésére,[69] a dezinformációs kampányokra, az online politikai propagandára és azok algoritmusok kihasználásával történő térhódítására, vagy éppen egyes demográfiai körök kirekesztésére egészségügyi információk kapcsán.[70] Amennyiben egy VLOPSE azonosít egy ilyen elfogultsági pontot, abban az esetben először értékelnie, majd csökkentenie kell az ebből eredő kockázatokat.[71] Utóbbira nevesített példa, hogy tesztelniük és ellenőrizniük kell a platformoknak az alkalmazott algoritmikus rendszereiket.[72] A DSA másik, főleg a kutatók számára előnyös rendelkezése, hogy a VLOPSE-knak, bár alapvetően szűk esetkörben, de „ismertetniük kell algoritmikus rendszereik – köztük ajánlórendszereik – tervezését, logikáját, működését és tesztelését" is,[73] azaz a DSA – helyesen – belátja, hogy a kutatókat szorosabban kell integrálni a VLOPSE-k kialakításából és használatából fakadó veszélyek azonosítása során.[74]

Az uniós szabályozás legfontosabb fejezete az algoritmikus szabályozás tekintetében ugyanakkor minden kétséget kizárólag a jelentős várakozás után végül 2024 nyarán kihirdetett mesterséges intelligenciáról szóló rendelet (*AI Act,* „AIA").[75] Bár az algoritmusok ritkán jelennek meg szövegszerűen a AIA-ben (összesen tíz említés), több cikk is alkalmazható az algoritmikus elfogultság jelenségére; a következőkben e szegmenseket ismertetjük. Mindenek előtt viszont egy szemantikai szempontból fontos előkérdést érdemesnek tartunk megemlíteni. Az angol nyelvű rendeletszövegben ugyanis az *„algorithmic bias"* nem jelenik meg, ugyanakkor a *„bias"* több ízben is szerepel. A magyar nyelvű fordítás ez utóbbit következetesen „torzításként" rögzíti, kissé szerencsétlen módon, hiszen a torzítás az angol szövegben konzekvensen a *„distort/distortion"* szóval jelöltetik, amelyre a magyar fordítás ugyancsak a torzítást használja.[76] Ennek említését csupán azért tartjuk relevánsnak, mert bár a következőkben citált cikkekben a torzítás szerepel *ad verbum*, jelen tanulmányban továbbra is az elfogultság kifejezéssel élünk.[77]

---

69 Ld. a *„shadowbanning"* jelenségét.

70 Raj M. Ratwani – Karey Sutton – Jessica E. Galarraga: Addressing AI Algorithmic Bias in Health Care. *JAMA,* Vol. 332., No. 13. (2024). https://doi.org/10.1001/jama.2024.13486

71 DSA 34–35. cikkek.

72 DSA 35. cikk (1) bekezdés d) pont.

73 DSA 40. cikk.

74 Anna Liesenfeld: The Legal Significance of Independent Research based on Article 40 DSA for the Management of Systemic Risks in the Digital Services Act. *European Journal of Risk Regulation,* (2024) 6–8. https://doi.org/10.1017/err.2024.61

75 Az Európai Parlament és a Tanács (EU) 2024/1689 rendelete (2024. június 13.) a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint a 300/2008/EK, a 167/2013/EU, a 168/2013/EU, az (EU) 2018/858, az (EU) 2018/1139 és az (EU) 2019/2144 rendelet, továbbá a 2014/90/EU, az (EU) 2016/797 és az (EU) 2020/1828 irányelv módosításáról (a mesterséges intelligenciáról szóló rendelet) (EGT-vonatkozású szöveg) PE/24/2024/REV/1 HL L, 2024/1689, 12.7.2024.

76 Ehhez ld. AIA (29) preambulumbekezdés.

77 Többek között azért is ragaszkodunk ehhez, mert a magyar fordítás számos szemantikailag nehezen értelmezhető terminust használ, így például nehezen interpretálható a „méltánytalan torzítás" (*unfair bias,* (27) preambulumbekezdés), amely jelentéstartalmát tekintve jóval inkább az olyan túlzó módon elfogult algoritmikus folyamatokat takarja, amelyek valamilyen módon negatív hatással járnak a fel-

Az elfogultság elsőként az AIA 27. preambulumbekezdésében jelenik meg, amely kiemeli, hogy az algoritmikus elfogultságot a sokszínűség, a méltányosság és a megkülönböztetésmentesség előmozdításával kell kezelni a mesterséges intelligencia rendszerekben. A hivatkozott preambulumbekezdés utal a 2019. április 8-án a mesterséges intelligenciával foglalkozó magas szintű uniós szakértői csoport által bemutatott megbízható mesterséges intelligenciára vonatkozó etikai iránymutatásokra is, amelyek a befogadó fejlődést és a tisztességtelen vagy diszkriminatív hatások elkerülését szorgalmazzák.[78] Az iránymutatások, ellentétben az AIA-val, lefektetik az elfogultság fogalmát, amely szerint a jelenség olyan tendenciákra vagy előítéletekre vonatkozik, amelyek (1) befolyásolhatják az kimeneteleket *(outputs)*, és (2) amelyek különböző forrásokból, például az adatgyűjtésből, a szabálytervezésből, a felhasználói interakcióból vagy korlátozott alkalmazási kontextusból erednek.[79] Az iránymutatás – helyesen – azt is aláhúzza, hogy míg az elfogultság „természetét" tekintve lehet szándékos vagy nem szándékos, és bizonyos esetekben akár előnyös is, az esetek többségében az algoritmusok torzító jellege diszkriminatív eredményekhez vezethez – ezt az iránymutatás tisztességtelen elfogultságnak *(unfair bias)* nevezi.[80] A tisztességtelen elfogultság elkerülése érdekében az iránymutatás hangsúlyozza az adekvát problémakezelés fontosságát, így például az adatokban meglévő (inherens) előítéletek és a diszkriminatív algoritmikus-tervezés elkerülését és orvoslását.[81] Ugyan az iránymutatás erősen absztraháltan fogalmaz, az „enyhítési stratégiák" közé sorolja az átlátható felügyelet biztosítását, valamint a fejlesztőcsapatok sokszínűségének előmozdítását.[82]

Az AIA 48. preambulumbekezdése aláhúzza, hogy a mesterséges intelligencia rendszerek (MI-rendszerek[83]) potenciálisan káros hatással lehetnek az alapvető jogokra, beleértve a megkülönböztetésmentességet, az egyenlőséget és a méltányosságot.[84] Bár az AIA kockázatalapú-struktúrájának részletes ismertetése meghaladná e dolgozat kereteit, aláhúzandó, hogy – akárcsak a DSA esetében – az MI-rendszereknél is differenciált rendszert alkalmaz a rendelet. Ennek értelmében az MI-rendszereket azok kockázata szerint osztályozza a jogszabály, ahol a tiltott, a nagy kockázatú és az általános, de rendszerszintű kockázatot jelentő MI-rendszerekre más szabályok vonatkoznak. Ugyan a kockázat-rendszer, amely a rendeletjavaslatban distinktív módon, négy részre volt különítve, jelentősen változott a kihirdetett szövegben,[85] az algoritmusok szerepe

használóra nézve (beleértve a méltányosság kérdését is). Azaz, tulajdonképpen az „előítéletességre", mintsem a torzításra utal.

78 A teljes dokumentum elérhető itt: European Commission: Ethics guidelines for trustworthy AI. 2019. április 8. https://tinyurl.com/2nyzyy6k

79 Uo. 38.

80 Uo.

81 Uo. 18.

82 Uo.

83 AIA 3. cikk 1. pont., vö. (28) prb.

84 Vö. A Biden-féle kormányzat kapcsán korábban említett *executive order*.

85 Claudio Novelli – Federico Casolari – Antonino Rotolo – Mariarosaria Taddeo – Luciano Floridi: Taking AI risks seriously: a new assessment model for the AI Act. *AI & Society,* Vol. 39., No. 5. (2023) 2493–2497. https://doi.org/10.1007/s00146-023-01723-z

és jelentősége a jogszabályban összességében – tartalmát tekintve – megegyezik a korábbi szövegváltozatokkal. E kontextusban a tiltott MI-rendszerek tekintetében az EU területén tilosak az olyan MI-gyakorlatok, így az olyan algoritmusok implementálása, amelyek súlyosan beavatkoznak és torzítják az emberi döntéshozatalt, kihasználják a felhasználó „sebezhetőségét", osztályozzák az embereket bizonyos kritériumok szerint (ez a passzus egyértelműen a kínai kreditrendszerre utal), bűnelkövetés megjóslásához és arcfelismerő adatbázisok létrehozásához használhatóak, természetes személyek érzelmeiből következtetéseket vonnak le,[86] biometrikus kategorizálást alkalmaznak, és „valós idejű" távoli biometrikus azonosító rendszereket használva a nyilvánosság számára hozzáférhető helyeken bűnüldözési célokat szolgálnak.[87] A biometrikus azonosítás tekintetében külön ki kell emelni a 32. preabulumbekezdést, amely tételesen utal az elfogultság kérdésére.[88] A bekezdés értelmében a valós idejű biometrikus azonosítás által létrejövő „torzított eredmények és diszkriminatív hatások különösen relevánsak az életkor, az etnikai és a faji hovatartozás, a nem vagy a fogyatékosságok tekintetében,"[89] így ezért is tiltott *(lex generalis)* az alkalmazása.[90]

A nagy kockázatú MI-rendszereknél, a DSA-hoz hasonló módon, az AIA is bevezet egy kifejezetten az érintett MI-gyakorlat kialakítására és működésére vonatkozó kockázatértékelési rendszert, amely az adott rendszer teljes életciklusa során megköveteli a kockázatok folyamatos azonosítását, elemzését és mérséklését, beleértve az algoritmikus elfogultságot is.[91] Bár expressis verbis az algoritmusokat nem említi, tartalmát tekintve az AIA 13. cikke, amely a nagy kockázatú rendszerek átláthatóságáról rendelkezik, nagy mértékben értelmezhető az algoritmusok transzparenciájára vonatkozó szabályként is. E körben az AIA megköveteli, hogy legyen világos és átfogó dokumentáció az érintett mesterséges intelligencia rendszer képességeinek és korlátainak magyarázatára, beleértve annak lehetőségét, hogy elfogult eredményeket produkáljon. Érdemes kiemelni e körben a (67) preambulumbekezdésben részleteiben tárgyalt „tanításhoz, validáláshoz és teszteléshez használt adatkészletek" kérdését. E körben az algoritmikus elfogultság kérdése az olyan adattömegek kapcsán merül fel, amelyek kifejezetten személyekre vagy csoportokra vonatkozó statisztikai adatokat tartalmaznak és amelyek ezáltal potenciálisan kirekesztőek lehetnek „bizonyos kiszolgáltatott csoportokkal" szemben. A kockázatok mérsékléséhez az ilyen adatkészleteknek rendelkeznie kell azokkal a jellemzőkkel és tulajdonságokkal, amelyek „azon sajátos földrajzi, kontextuális, magatartási vagy funkcionális környezethez kapcsolódnak, amelyben az MI-rendszert használni szándékozzák." A vonatkozó 10. cikk pedig konkrétan lefek-

---

86 Ehhez bővebben ld. AIA 44. preambulumbekezdés, amely külön kihangsúlyozza az általánosíthatóság problémáját.

87 AIA 5. cikk (1) bekezdés. Fontos ugyanakkor megemlíteni, hogy egyes tiltott MI-gyakorlatok esetében, így például a bűnüldözési célokra használható rendszerek tekintetében fennállhatnak kivételes esetek, amikor e rendszerek használata megengedett.

88 „[…] természetes személyek távoli biometrikus azonosítására szolgáló MI-rendszerek technikai pontatlansága torzított eredményekhez vezethet, és diszkriminatív hatásokat eredményezhet"

89 Uo.

90 Ugyancsak ld. AIA (54) prb.

91 AIA 9. cikk.

teti, hogy az ilyen nagy kockázatú MI-rendszereknek (magyarán olyanoknak, amelyek a tanító, a validálási és a tesztadatkészletekkel dolgoznak), kötelezően szükséges minden olyan lehetséges kockázat kivizsgálása, amelyek az alapvető jogokat sérthetik, megkülönböztetéshez vezetnek, illetve ha bármilyen tekintetben hatnak (azaz nemcsak negatív, hanem akár semlegesen vagy pozitívan is) az érintett személyek egészségére és biztonságára.[92] E körben hangsúlyos a kivizsgálás akkor, ha „az adatok kimenetei befolyásolják a jövőbeli műveletek bemeneteit," azaz például olyan esetekben ahol korábbai adatokra építve tesz majd az MI később javaslatokat; hipotetikusan például olyan városrészeken javasol magasabb rendőri jelenlétet, ahol korábban több bűnözés volt. A kivizsgálás mellett az ilyen MI-rendszereknél szükséges preventív és enyhítő intézkedéseket is bevezetni.[93] Ezen túlmenően, a 10. cikk 5. pontja kiemeli, hogy a magas kockázatot jelentő MI-rendszerek szolgáltatói a személyes adatok különleges kategóriáit (pl. faji, egészségügyi vagy politikai meggyőződésre utaló adatokat) csak és kizárólag akkor kezelhetik, ha az elfogultságok felderítéséhez és korrigálásához feltétlenül szükségesek továbbá, ha az elfogultságok felderítése semmilyen más módon nem lehetséges.[94] Az ilyen adatfeldolgozásnak szigorú biztosítékokat kell követnie, beleértve az álnevesítés, a korlátozott hozzáférés, a biztonsági ellenőrzések, a harmadik felekkel való megosztás tilalma és az időben történő törlés.[95] Továbbá részletes dokumentációval kell igazolni azt is, hogy az érzékeny adatok felhasználása miért volt szükséges, illetve bizonyítani kell az uniós adatvédelmi jogszabályoknak való megfelelést. Kiemelendők még a magas kockázatú MI-rendszereknél a 14. cikk rendelkezései, amelyek aláhúzzák, hogy az ilyen rendszereknek hatékony emberi felügyeletet kell biztosítaniuk az egészséget, a biztonságot és az alapvető jogokat érintő kockázatok minimalizálása érdekében. Az algoritmikus elfogultság kontextusában kulcsfontosságú a 4. pont b), amely kifejezetten az automatizálási torzítás kérdését érinti.[96] E körben a rendelet megköveteli, hogy a felügyeleti mechanizmusok segítsék a felhasználókat abban, hogy tudatában legyenek ennek a kockázatnak, lényegében annak biztosítékaként, hogy az emberek képesek legyenek megkérdőjelezni vagy akár felülbírálni a mesterséges intelligencia döntéseit. A tervezés és az elfogultság kapcsán a 15. cikk szolgált támpontként. E cikk kiemeli, hogy a nagy kockázatú MI rendszereket úgy kell megtervezni, hogy minimalizálják az elfogult visszacsatolási hurkokat,[97] különösen azokban a rendszerekben, amelyek a telepítés után is folytatják a tanulást. Ez magában foglalja olyan technikai és szervezeti intézkedések végrehajtását is, amelyek megaka-

92 AIA 10. cikk 2. pont f) bekezdés.

93 AIAI 10. cikk 2. pont g) bekezdés.

94 Részletes elemzéshez ld. Marvin van Bekkum: Using sensitive data to de-bias AI systems: Article 10(5) of the EU AI act. *Computer Law & Security Review,* Vol. 56. (2025) https://doi.org/10.1016/j.clsr.2025.106115

95 Ld. Uo. AIA 5.pont (a)–(f).

96 Ez egy tendenciát jelent, amelynek lényege, hogy az emberek túlzottan is az MI által létrehozott eredményekre támaszkodnak.

97 Általánosságban véve azt értjük rajta, amikor egy MI-rendszer kimenetei a jövőbeli tanulás vagy döntéshozatal bemeneteként szolgálnak.

dályozzák, hogy az elfogult kimenetek negatívan befolyásolják a jövőbeli bemeneteket, és biztosítják az ilyen kockázatok aktív csökkentését is.

5. Az algoritmikus elfogultság mint lernéi Hüdra és a jövő szabályozási kérdései

Mitológiai hasonlattal élve, az MI, különösen az algoritmusok szabályozása, hasonlít Héraklész párbajára a lernéi Hüdrával –minden szabályozási próbálkozásra két újabb, addig ismeretlen vagy jogi eszközökkel kezelhetetlen probléma kerül elő. Az egyik legfontosabb, továbbra is megoldásra váró polémia az algoritmusok és adatkészletek kialakításában rejlő elfogultság. Kétségtelen, hogy akár a DSA, akár az AIA által bevezetett átláthatósági szabályok fontos lépések, ugyanakkor továbbra is fennáll az algoritmusok kialakításakor megjelenő feketedoboz*(„black box")*-hatás, amely tulajdonképpen az MI-rendszerekbe inherensen „kódolt" átláthatatlanságot jelenti.[98] A feketedobozok átláthatatlansága lehetetlenné teszi a probléma gyökerének megtalálását, azaz a elfogultságok azonosítását és orvoslását, még akkor is, ha a rendeletek által előírt dokumentációs kötelezettség teljesülne.

További akadály lehet a végrehajtás. Bár különösen az AIA lefektet végrehajtási szabályokat, kérdéses, hogy biztosítható-e megfelelő forrással és leginkább szakértelemmel a jogszabály megfelelő betartatása. E tekintetben kifejezetten aggasztó az MI-hivatal, amelynek célja, hogy hozzájáruljon „az MI-rendszerek és az általános célú MI-rendszerek, valamint az MI-irányítás végrehajtásához, nyomon követéséhez és felügyeletéhez".[99] Az elmúlt egy évben elsősorban adminisztratív eredményeket tud felmutatni, míg a Hivatal munkáját támogató MI-testület eddig összesen kétszer ülésezett, érdemi eredmény nélkül.[100] Az erőforrások hiánya és a lassú bürokratikus folyamatok még inkább láthatóak nemzeti szinten, ugyanis a tagországok többségében egyelőre nincs kifejezetten az MI-re és annak ellenőrzésére szakosodott hatóság. Az erőforrás és szakértelem e körben messzemenőleg túlmutat az AIA ismeretén; súlyos erőforráshiány van mind technológiai, mind akár MI-etikai szakértőből.[101]

Talán még nagyobb, és a jogi szabályozáson túlmutató probléma az algoritmusok szabályozásának összehangolatlansága. Ahogy az ENSZ Főtitkára által összeállított MI-vel foglalkozó szakértői panel is jelezte 2024 őszén: „megdönthetetlen" szükséglet egy globális szabályozás kialakítása és nem szabad engedni, hogy kizárólag a piaci erők diktálják az MI-fejlesztést és a szabályozás korlátait.[102] A több, mint száz oldalas riport arra az aggodalomra is kitér, hogy megfelelő, globális kormányzás nélkül a mes-

---

98 Bartosz Brożek – Michał Furman – Marek Jakubiec – Bartłomiej Kucharzyk: The black box problem revisited. Real and imaginary challenges for automated legal decision making. *Artificial Intelligence and Law,* Vol. 32., No. 2. (2023) 427–428. https://doi.org/10.1007/s10506-023-09356-9

99 AIA 3. cikk 47. pont.

100 Ld. European AI Office. *EU*, 2024. https://digital-strategy.ec.europa.eu/en/policies/ai-office

101 Az egyik legnagyobb számítástechikai újság, a ComputerWeekly felmérése szerint az európai IT-ban dolgozók 40%-ának jelentős hiányosságai vannak az AI kapcsán. Clare McDonald: AI and cyber skills worryingly lacking, say business leaders. *ComputerWeekly,* 2024. július 11. https://tinyurl.com/27kn7p96

102 Governing AI for Humanity. *UN Final Report,* 2024. szeptember. 7–8.

terséges intelligencia előnyei csak néhány kiválasztott nemzetre és szervezetre korlátozódhatnak, amely tovább súlyosbíthatja az eddig is fennálló digitális megosztottságot és az egyenlőtlenségeket.[103] A harmonizáltság idealisztikus, vagy talán romanticizált képe viszont a gyakorlatban igen képlékeny talajon áll. Sajnálatos módon ugyanis a fentebb említett anyag nem tér ki arra, hogy az algoritmikus elfogultság nem csak és kizárólag technológiai kérdés; sőt, számos olyan társadalmi és kulturális elemet is hordoz magában (akár közvetve vagy közvetlenül), amelyek a globális szabályozással nem tűnnek feloldhatónak. A szabályozási harmonizáció ugyanis jelen esetben egy globális „társadalmi elv" elfogadását is jelenti, amely az emberi jogok és a méltányosság értelmezése tekintetében is közös értékeket feltételez – helytelenül és alaptalanul. Kritikai szemszögből vizsgálva a harmonizáció kérdését megállapítható az ENSZ-jelentés kapcsán, hogy a javasolt normakezdeményezések a nyugati, demokratikus elvekből erednek, amelyek nem feltétlenül elfogadottak vagy érvényesülnek általánosan – gondolhatunk itt olyan régiók országaira, ahol a nők és marginalizált közösségek védelme folyamatosan sérül. Így a diszkrimináció csökkentésére törekvő nemzetközi normák végérvényesen egy általános jogalkotói és társadalmi attitűdöt is közvetítenek, amelyek bizonyos szocio-kulturális kontextusokban ütközhetnek az uralkodó intézményi vagy ideológiai struktúrákkal. Végül érdemes a jövőben azt is figyelembe venni, különösen a fentiek fényében, hogy az algoritmikus elfogultság globalizált szabályozása egyáltalán szolgálhat-e, sőt, kellene-e szolgálnia az etikai értékek technikai szabályozásnak álcázott exportjának eszközeként. Ennek vizsgálatához fontos lehet a normatív imperalizmus elkerülése,[104] amely a helyi társadalmi-politikai realitásoktól függetlenül a nyugati prioritásokat beágyazza a határokon átnyúlóan működő algoritmusokba.

6. Következtetések

A fenti metaforát folytatva, Héraklésznak szüksége volt Iolaoszra, aki (bár némi isteni segítséggel) megtalálta a megfelelő fogást a Hüdrán, s így segítségével a thébai hős legyőzte a szörnyet. Az algoritmikus elfogultság szabályozásánál, egyetértve az ENSZ riportjával, Iolaosz nem egy jogszabály, egy újszerű rendelet vagy akár az MI-fejlesztők saját szabályozása, hanem jóval inkább egy globális szintű, átfogó, kooperatív szabályozás lehet. Ez a kooperatív munka sokkal többet takar, mint egyes államok közös erőfeszítését; be kell vonni a civil szervezeteket, a kutatókat, a különböző érdekképviseleti szervezeteket, az MI-ben aktívan részt vevő gazdasági szereplőket és természetesen hangot kell adni a marginalizált csoportoknak is, akiket a leginkább érint az algoritmusok kódolt kirekesztése.

Jelen tanulmányban az algoritmikus elfogultság kérdését vizsgáltuk, különös tekintettel annak fogalmi alapjaira és szabályozás jelenlegi helyzetére. A dolgozatban ismertettük a legfontosabb jogi irányokat, és kiemelt hangsúlyt fektettünk az európai (uniós) szabályok bemutatására. A kutatás kiemelt eredménye – egyben fontos felhívása –,

---

103 Uo. 27.

104 Ld. Julian Pänke: The Fallout of the EU's Normative Imperialism in the Eastern Neighborhood. *Problems of Post-Communism,* Vol. 62., No. 6. (2015) 350–363. https://doi.org/10.1080/10758216.2015.1093773

hogy sürgősen szükség van egy harmonizált, globális szabályozási keretre. A jelenlegi regionális és nemzeti erőfeszítések ugyanis – bár dicséretesek – nem elegendőek a mesterséges intelligencia fejlesztésének és alkalmazásának határokon átnyúló jellegének kezelésére. A jövőbeni kutatásokra javasolt az algoritmikus elfogultsággal kapcsolatos technikai megoldások bemutatása, a jogi kérdések tekintetében pedig fontos eredményekkel szolgálhatnak majd a hazai irányokat előrevetítő vagy akár az EU-n és Egyesült Államokon kívüli szabályozási kezdeményezések és megoldások bemutatása.